DOCUMENT RESUME

ED 331 892 TM 016 500

AUTHOR Holdzkom, David; Stacey, Dennis
TITLE The Limits of Generic Skills Evaluation: Using the NCTPAS To Evaluate Special Teaching Assignments.
PUB DATE Apr 91
NOTE 29p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, April 3-7, 1991).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Elementary School Teachers; Elementary Secondary Education; Evaluation Methods; Kindergarten; *Secondary School Teachers; Special Education; *Specialists; Teacher Effectiveness; *Teacher Evaluation; Teaching Skills; Vocational Education
IDENTIFIERS *Generic Skills; *North Carolina Teacher Performance Assess System

ABSTRACT

In the 1985-86 school year, North Carolina implemented a uniform system of teacher performance assessment--the North Carolina Teacher Performance Assessment System (NCTPAS). The evaluation results for teachers with special assignments were examined, since many teachers felt that their special areas did not lend themselves to successful classroom observation. The evaluation system uses the observations of trained observers who rate the teacher on 28 identified teaching practices and also prepare a narrative summary of teacher performance. It was hypothesized that ratings of teachers in special assignment areas would not differ from ratings for other types of classroom teaching. These assignment areas were: (1) kindergarten; (2) combination classes; (3) classes for exceptional children and vocational education; and (4) secondary classes in specific subject areas. Evaluation data were gathered for over 5,000 teachers in grades 1 through 5 in both 1988 and 1989. Comparisons indicate that the generic skills evaluation system did not discriminate against any class of teachers. It is also apparent that evaluators could use the system fairly and that teachers were able to improve their skills based on feedback from the observation system. Nine tables of evaluation data and a 27-item list of references are included. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED331892

The Limits of Generic Skills Evaluation: Using the NCTPAS to Evaluate Special Teaching Assignments

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

DAVID HOLDZKOM

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

by

David Holdzkom
Chief Consultant
LEA Personnel Services

and

Dennis Stacey
Chief Consultant
Student Support Services

North Carolina Department of Public Instruction
Raleigh, North Carolina

BEST COPY AVAILABLE

TM016500

A paper presented at the annual meeting of AERA, Chicago, Illinois, April 1991

During school year 1985-86, North Carolina implemented a uniform system of performance assessment of the 60,000+ teachers employed in the State. The evaluation system was predicated on the results of the "teacher effects" research, relied on classroom observation as the primary data source, and had both formative and summative applications. Since implementation, the evaluation system has been carefully and systematically studied. Two groups of third-party evaluators have reported on the evaluation system and its application in 16 of the 134 school districts. The distribution of rating points has been studied annually. The changes in student achievement during the first four years of implementation have been analyzed, as have the changes in teachers' and evaluators' skills. This report continues that tradition of looking for effects of evaluation on teachers' behavior. Here, we will examine the evaluation results of identifiable "minorities" of teachers: those who teach elementary combination classes, kindergarten teachers, vocational educators, special educators, "specials" (P.E., music, and art), as well as secondary school teachers who teach a variety of subjects.

Background

Historically, education in North Carolina has been a highly centralized function. While local boards of education enjoy a fair degree of autonomy for operational decisions, the State has paid the lion's share of the costs. In the current fiscal year, the State will pay 67% of all educational expenses and 73% of all

salary costs. However, the State has a fairly complex formula for allotments, so that the percentage of costs of any given school district will vary from these averages. For example, the State provides $100 per year per teacher for staff development, but pays the full amount of teachers' salaries, based on the state's uniform salary schedule. Thus, poor districts have an incentive to hire the best teachers they can find, not the cheapest, but the percentage of costs borne by the state is relatively high.

Teacher performance assessment is another example of the partnership between the State and local boards. North Carolina General Statute 115C-326 requires that the State Board of Education adopt a uniform system of teacher evaluation to be implemented by the local boards. Upon adoption of the statute in 1979, the State Department of Public Instruction (DPI) developed an evaluation process that rested on criteria of teacher effectiveness as determined by surveys of teachers and principals. This "consensus" instrument was viewed as a beginning in the domain of teacher assessment and local districts were granted wide latitude in its implementation.

In 1982, DPI granted a contract to the Group for Effective Teaching at the University of North Carolina - Chapel Hill to review the research literature on teacher effectiveness. The Group identified several criteria to which research studies were subjected before being included in the data base. The studies had to be empirical, to identify skills that were observable, alterable, and present in multiple grades/subject matters. Finally, the studies had to demonstrate a relationship between the

skills identified and improvements in student achievement and/or increased time on task (The Group, 1983). In all, more than 100 different studies that met all the criteria were identified. Analysis of these studies yielded a list of five teaching functions: management of instructional time, management of student behavior, instructional presentation, instructional monitoring, and instructional feedback. Each of these was operationally defined by "indicators" that specified teaching practices: e.g., the teacher uses demonstrations to illustrate concepts; the teacher circulates to monitor students' work; the teacher provides specific feedback on students' in-class work, etc. In all, the five functions encompassed 28 indicators. These, then, become the basis of the observation system developed for the evaluation procedure (Holdzkom, 1987). However, before the development of the procedures, the Group's work was reviewed by a panel of experts and was tested in a field test.

Based on the results of the field test and the panel's advice, procedures were specified (The Group, 1985a and 1985b). Each teacher would be observed by trained observers who would be in the classroom for a full instructional period. The observer's view was shaped by the teacher's demonstrations of the 28 practices, but data were collected using a script tape or field notes that later were codified by the observer. Then, a narrative summary of the teacher's performance (analyzed at the function level) was prepared and discussed with the teacher in a conference. For summative purposes, multiple observations were conducted and a numerical rating (on a scale of 1-6) would be derived for each function.

The task of training both evaluators and teachers to be evaluated fell to DPI staff. A 30-hour training program was developed and delivered, via turn-key training, to over 60,000 teachers between July 1985 and December 1987. In addition, a 24-hour training program for evaluators was delivered to principals and others in administrative roles. This training culminated in a performance test for evaluators that established the ability to conduct evaluation reliably. (Subsequently, booster training activities were developed and implemented about every 18 months.)

Simultaneous with the implementation of the NCTPAS, a second, related project was launched by DPI. A Career Development Program (CDP) pilot was implemented in 16 school districts. This program, primarily designed to test various career incentives, was founded on the results of the NCTPAS. Thus, a significant effort was made in the 16 districts to implement the evaluation process in conditions that duplicated as closely as possible the optimum. Because virtually all teachers in the districts elected to participate in CDP, generalizations about implementation effects could safely be made.

After 18 months (two evaluation cycles), a study of evaluation implementation was conducted in 35 school districts (Stacey, Holdzkom, and Kuligowski, 1989). Among other things, this study revealed that most of the teachers (N=2732) and evaluators (N = 639) agreed that the criteria were appropriate and that the procedures for evaluation were fair. Moreover, a third-party evaluation, conducted by a panel of nationally-renowned experts, found that the evaluation instrument "was a quality instrument, one

that is highly suited to its purposes" (Brandt et al., 1988). In addition, a second third-party evaluation of the CDP pilot found a positive sentiment among teachers that their evaluations were fair and appropriate (Furtwengler, 1988).

The annual implementation of performance appraisal in the 16 CDP districts was reported to the State Board of Education. It was found, in 1987, that teachers earning acceptable or better ratings (3, 4, 5 or 6) distributed into perfect bell-shaped curves on the five observable functions (NCDPI, 1987). This was especially important because it showed that--at the state-level--the instrument could discriminate among teachers without forcing artificial quotas at the district level. Moreover, when, in 1988 and 1989, the curves began skewing to the right, indicating higher average ratings, it was concluded that teachers' performance had improved, probably as a result of incentives, feedback, and staff development. (Holdzkom, Kuligowski, and Stacey, 1989).

Finally, an analysis of student performance found that in Career Development pilot districts, student achievement exceeded that of a matched sample of students in other districts (Holdzkom, Kuligowski, and Stacey, 1990). Clearly, a link existed between improving teacher performance and increased student achievement. It should, however, be observed that student achievement gains were not used in North Carolina as a teacher evaluation criterion, but as a program evaluation strategy (Brandt, 1990).

Despite the congruity of all these findings, however, at least some teachers felt, perhaps intuitively, that a system of generic teacher evaluation would result in negative effects on certain

classes of teachers. For example, educators of exceptional children were fearful that the unpredictable behavior of their students might result in observers "marking down" teachers for poor student behavior management. Other groups of teachers--notably art teachers, vocational educators, and some exceptional children's teachers--voiced concern that their teaching "didn't observe well". Presumably, the individualized nature of instruction or the emphasis on laboratory work would be misconstrued by evaluators who were widely (if incorrectly) perceived to be interested only in a Hunteresque model of instruction calling for a specific sequence of instruction with a clearly didactic role for the teacher. Similarly, kindergarten teachers, French teachers, chemistry teachers and others asserted that the evaluators' lack of understanding of either the students or the subject matter or both would render the evaluation invalid. Indeed, some support for this position, at least at the level of theoretical discussion, appears to exist. Susan Stodolsky argues cogently for the need for evaluator sensitivity to the wide range of teacher behaviors that are varied depending upon instructional format, pacing, and cognitive level (Stodolsky, 1984). Moreover, those familiar with Shulman's work will recognize the connection between his activities and the fears expressed by some teachers.

Despite our confidence in the NCTPAS and the evaluators implementing it, we felt that these concerns of teachers merited our attention. Accordingly, we formulated three main hypotheses. These were:

Hypothesis Ia: Kindergarten teachers would receive

performance ratings at the same quality level as all elementary teachers.

Hypothesis 1b: The ratings of teachers of combination classes (e.g. K-1, 1-2, etc.) would be on a par with ratings of teachers assigned classes at the component, single grade levels (e.g., kindergarten, first grade, etc.).

Hypothesis II: Ratings of teachers of exceptional children, vocational educators, and of "special" teachers would not be different from ratings of all other teachers.

Hypothesis III: Ratings of secondary school teachers would not vary as a function of subject area from ratings of all teachers.

Method

Evaluation ratings on each of the eight functions of the NCTPAS (Figure 1) for every teacher in the 16 Career Development pilot districts were reported to DPI annually. Individual teacher data provided identifier information including grade or subject of primary assignment, school, years of experience, and CDP status (beginner, Level I, Level II). For this study, data were aggregated on the basis of primary job assignment ("class") at the state level. Scoring means for each function were calculated, as were standard deviations. These could then be compared to those for other classes and for all participating teachers. Data presented here were collected at the

1. Major Function: *Management of Instructional Time*
   - 1.1 Teacher has materials, supplies, and equipment ready at the start of the lesson or instructional activity.
   - 1.2 Teacher gets the class started quickly.
   - 1.3 Teacher gets students on task quickly at the beginning of each lesson or instructional activity.
   - 1.4 Teacher maintains a high level of student time-on-task.

2. Major Function: *Management of Student Behavior*
   - 2.1 Teacher has established a set of rules and procedures that govern the handling of routine administrative matters.
   - 2.2 Teacher has established a set of rules and procedures that govern student verbal participation and talk during different types of activities—whole-class instruction, small-group instruction, and so on.
   - 2.3 Teacher has established a set of rules and procedures that govern student movement in the classroom during different types of instructional activities.
   - 2.4 Teacher frequently monitors the behavior of all students during whole-class, small-group, and seatwork activities and during transitions between instructional activities.
   - 2.5 Teacher stops inappropriate behavior promptly and consistently, yet maintains the dignity of the student.

3. Major Function: *Instructional Presentation*
   - 3.1 Teacher begins lesson or instructional activity with a review of previous material.
   - 3.2 Teacher introduces the lesson or instructional activity and specifies learning objectives when appropriate.
   - 3.3 Teacher speaks fluently and precisely.
   - 3.4 Teacher presents the lesson or instructional activity using concepts and language understandable to the students.
   - 3.5 Teacher provides relevant examples and demonstrations to illustrate concepts and skills.
   - 3.6 Teacher assigns tasks that students handle with a high rate of success.
   - 3.7 Teacher asks appropriate levels of questions that students handle with a high rate of success.
   - 3.8 Teacher conducts lesson or instructional activity at a brisk pace, slowing presentations when necessary for student understanding but avoiding unnecessary slowdowns.
   - 3.9 Teacher makes transitions between lessons and between instructional activities within lessons efficiently and smoothly.
   - 3.10 Teacher makes sure that the assignment is clear.
   - 3.11 Teacher summarizes the main point(s) of the lesson at the end of the lesson or instructional activity.

4. Major Function: *Instructional Monitoring of Student Performance*
   - 4.1 Teacher maintains clear, firm, and reasonable work standards and due dates.
   - 4.2 Teacher circulates during classwork to check all students' performance.
   - 4.3 Teacher routinely uses oral, written, and other work products to check student progress.
   - 4.4 Teacher poses questions clearly and one at a time.

5. Major Function: *Instructional Feedback*
   - 5.1 Teacher provides feedback on the correctness or incorrectness of in-class work to encourage student growth.
   - 5.2 Teacher regularly provides prompt feedback on assigned out-of-class work.
   - 5.3 Teacher affirms a correct oral response appropriately, and moves on.
   - 5.4 Teacher provides sustaining feedback after an incorrect response or no response by probing, repeating the question, giving a clue, or allowing more time.

6. Major Function: *Facilitating Instruction*
   - 6.1 Teacher has an instructional plan that is compatible with the school and system-wide curricular goals.
   - 6.2 Teacher uses diagnostic information obtained from tests and other assessment procedures to develop and revise objectives and/or tasks.
   - 6.3 Teacher maintains accurate records to document student performance.
   - 6.4 Teacher has instructional plan that matches/aligns objectives, learning strategies, assessment, and student needs at the appropriate level of difficulty.
   - 6.5 Teacher uses available human and material resources to support the instructional program.

7. Major Function: *Communicating Within the Educational Environment*
   - 7.1 Teacher treats all students in a fair and equitable manner.
   - 7.2 Teacher interacts effectively with students, co-workers, parents, and community.

8. Major Function: *Performing Non-Instructional Duties*
   - 8.1 Teacher carries out non-instructional duties as assigned and/or as need is perceived.
   - 8.2 Teacher adheres to established laws, policies, rules, and regulations.
   - 8.3 Teacher follows a plan for professional development and demonstrates evidence of growth.

Fig. 1. Teaching Functions and Practices

BEST COPY AVAILABLE

end of school years 1987-88 and 1988-89, the final two years of the four-year Career Development Pilot Program.

Results

Table Ia displays the mean scores on each of the first five functions of the NCTPAI broken down by grade-level for six elementary school grades, for 1988 and 1989. Using all K-5 teachers as a bench mark, it is clear that kindergarten teachers received higher average scores on Function 1 (Time Management); Function 3 (Instructional Presentation); Function 4 (Instructional Monitoring); and Function 5 (Instruction Feedback) than the average for all elementary school teachers in 1988. In 1989, this trend continued. Moreover, kindergarten teachers improved their ratings in Function 2 up to the average score for all K-5 teachers. Indeed, the table makes clear that in time management and instructional feedback, kindergarten teachers outperformed teachers at any other grade in both years.

Table Ib displays the mean scores on each of the first five functions for teachers teaching combination classes (K-1, 1-2, 2-3, 3-4, and 4-5). Since assignment of pupils to classes is done by school principals who work within local guidelines, it is impossible to state how combinations are formed. We do not know what percentage of classes, for example, are made up of high-achieving kindergarten/low-achieving first graders, or high kindergartners with high firsts, etc. However, it is clear from Table 1b that the teacher evaluation ratings are less stable for those teachers than for any other group. For example, while scores

TABLE Ia
Mean Scores by Function for Teachers
by Grade Taught, 1988 and 1989

	K	1	2	3	4	5	All K-5 Teachers (incl.comb.)
1988 Func. 1	4.67	4.57	4.55	4.66	4.42	4.60	4.58
2	4.51	4.49	4.52	4.68	4.51	4.68	4.56
3	4.66	4.53	4.53	4.58	4.35	4.61	4.55
4	4.61	4.51	4.55	4.63	4.41	4.58	4.55
5	4.60	4.50	4.47	4.49	4.33	4.53	4.49
	(N=317)	(N=294)	(N=293)	(N=279)	(N=278)	(N=271)	(N=906)
1989 Func. 1	4.88	4.77	4.66	4.75	4.68	4.80	4.76
2	4.74	4.66	4.67	4.78	4.69	4.87	4.74
3	4.88	4.70	4.61	4.73	4.53	4.78	4.71
4	4.71	4.68	4.60	4.72	4.59	4.72	4.67
5	4.73	4.64	4.57	4.63	4.55	4.67	4.61
	(N=305)	(N=310)	(N=292)	(N=298)	(N=270)	(N=299)	(N=1774)

TABLE Ib
Mean Scores by Function for Teachers of Combination Classes, 1988 and 1989

	K-1	1-2	2-3	3-4	4-5	All Combination Teachers*	All K-5 Teachers
1988 Function 1	4.63	4.96	4.36	4.86	4.56	4.67	4.58
2	4.27	4.88	4.48	4.79	4.67	4.56	4.56
3	4.44	4.84	4.39	4.68	4.56	4.57	4.55
4	4.49	4.92	4.55	4.71	4.52	4.60	4.55
5	4.54	4.72	4.52	4.61	4.56	4.59	4.49
	(N=41)	(N=25)	(N=33)	(N=28)	(N=27)	(N=172)	(N=1906)
1989 Function 1	5.07	4.52	4.82	4.68	4.67	4.81	4.76
2	5.14	4.64	4.82	4.94	4.51	4.74	4.74
3	4.86	4.76	4.76	4.81	4.50	4.71	4.71
4	4.93	4.52	4.79	4.84	4.59	4.73	4.67
5	4.86	4.64	4.65	4.81	4.53	4.72	4.61
	(N=28)	(N=25)	(N=34)	(N=31)	(N=70)	(N=320)	(N=1774)

*Includes combinations other than K-1, 1-2, 2-3, 3-4, and 4-5.

tend to rise from year to year, mean scores for 1st and 2nd combination teachers are uniformly lower in 1989 than they were in 1988. This same effect is seen for teachers of 3-4 combinations (Function 1) and 4-5 combinations (Function 2, 3, and 5), but nowhere else. Overall, combination teachers as a group outscored non-combination teachers on every function, except Function 2 (Behavior management), where a tie is recorded both years, and Function 3 in 1989 only.

However, at the class-level, some interesting patterns are seen, as displayed in Table Ic. This table compares mean ratings of combination teachers to ratings of all teachers of the constituent grades, with a + representing a case in which the single grade teachers' mean rating is higher than the combination class teachers' rating. Thus, in 1988, teachers of kindergarten only classes outscored K-1 combination teachers on every function. The same was true for fifth grade teachers, except for Function 5. The reverse was true for kindergarten teachers, however in 1989. It should be noted that two years may not provide a sufficient base to establish trends, and that the comparisons occur in a context of general improvement, with all means increasing. At the least, we have established that children assigned to combination classes do not necessarily receive instruction from less skilled teachers than children not so assigned. It is, however, interesting to observe that fifth grade only teachers outperform fifth grade combination teachers consistently.

TABLE Ic
Relationship of Mean Ratings of Single Class Teachers to Combination Class Teachers, 1988 and 1989

1988

FUNCTION	K	1		2		3		4		5
1	+	-	-	-	+	-	-	-	-	+
2	+	+	-	-	+	-	-	-	-	+
3	+	+	-	-	+	-	-	-	-	+
4	+	+	-	-	-	-	-	-	-	+
5	+	-	-	-	-	+	-	-	-	-

1989

FUNCTION	K	1		2		3		4		5
1	-	-	+	+	-	-	+	-	+	+
2	-	-	+	+	-	-	-	-	+	+
3	+	-	-	-	-	-	-	-	+	+
4	-	-	+	+	-	-	-	-	-	+
5	-	-	-	+	-	-	-	-	+	+

Taken together, these data support both parts of our first hypothesis: no systematic discrimination against either kindergarten or combination class teachers is associated with the NCTPAS.

Hypothesis II focuses on special education teachers, elementary teachers of art, music, dance, and physical education ("specials"), secondary school vocational educators, elementary basic skills remedial teachers, and secondary school teachers who teach two or more unrelated subjects. Table IIa presents the annual mean scores for these groups on the first five functions for 1988 and 1989. Special educators are sub-divided into those based in elementary schools and secondary schools. The most interesting fact shown here probably is that for all of these classes of teachers, the Function 3 (Instructional Presentation) mean rating is below the general mean for both 1988 and 1989 (except for "specials" in 1989). This could reflect a lack of awareness on the evaluators' part about what is being taught by the teacher in these classes. However, another possible explanation is that these teachers, as a group, do less "teaching" or do it less well than other teachers. While secondary level special educators received mean scores higher than the general mean in all other functions both years, elementary special educators and vocational educators and remedial specialists declined from 1988 to 1989 as compared to the general mean, and "special" teachers generally improved against the general mean. While all classes of teachers improved absolutely from 1988 to 1989, high school teachers of two or more unrelated subjects started and finished below the general mean.

TABLE IIa
Mean Ratings For Selected Groups
of Teachers, 1988 and 1989

1988

FUNCTION	All Teachers	Special Ed. (Elementary)	Special Ed. (Secondary)	"Specials" (Elem.)	Voc. Ed. (Secondary)
1	4.56	4.57	4.68	4.44	4.58
2	4.56	4.64	4.81	4.31	4.59
3	4.51	4.44	4.49	4.46	4.45
4	4.54	4.64	4.73	4.42	4.59
5	4.51	4.70	4.78	4.39	4.53
	(N=5835)	(N=414)	(N=103)	(N=304)	(N=493)

Remedial (Elementary)	2 or More (Secondary)
4.59	4.27
4.58	4.40
4.42	4.23
4.60	4.17
4.56	4.37
(N=166)	(N=30)

TABLE IIa (CONT'D)
Mean Ratings For Selected Groups
of Teachers, 1988 and 1989

1989

FUNCTION	All Teachers	Special Ed. (Elementary)	Special Ed. (Secondary)	"Specials" (Elem.)	Voc. Ed. (Secondary)
1	4.71	4.68	4.76	4.77	4.68
2	4.71	4.70	4.85	4.56	4.71
3	4.65	4.58	4.52	4.71	4.54
4	4.66	4.72	4.80	4.67	4.67
5	4.62	4.75	4.77	4.66	4.59
	(N=5915)	(N=234)	(N=88)	(N=145)	(N=476)

Remedial (Elementary)	2 or More (Secondary)
4.71	4.38
4.59	4.43
4.41	4.14
4.57	4.29
4.51	4.38
(N=70)	(N=21)

TABLE IIb
Comparison of Selected Group Means
to General Mean, 1988 and 1989

1988

Function	Spec. Ed. (Elem.)	Spec. Ed. (Sec.)	"Spec."	Voc. Ed.	Remedial	2 or more
1	+	+	-	+	+	-
2	+	+	-	+	+	-
3	-	-	-	-	-	-
4	+	+	-	+	+	-
5	+	+	-	+	+	-

1989

Function	Spec. Ed. (Elem.	Spec. Ed. (Sec.)	"Spec."	Voc. Ed.	Remedial	2 or more
1	-	+	+	-	-	-
2	-	+	-	-	-	-
3	-	-	+	-	-	-
4	+	+	+	+	-	-
5	+	+	+	-	-	-

Improvement from 1988 to 1989 (Group to self)

Spec. Ed. (Elem.)	Spec. Ed. (Sec.)	Specials	Voc. Ed.	Remedial	2 or More
+	+	+	+	+	+
+	+	+	+	+	+
+	+	+	+	-	-
+	-	+	+	-	+
+	+	+	+	-	+

(The small number of teachers in this class should not be overlooked.) Table IIb compares the mean for each function for each class with the general mean. A + indicates that the class mean exceeds the general mean. The lower part of the table compares the class performance in 1988 with 1989, with a + indicating improvement over time. From these data, it would appear that Hypothesis II is upheld.

Hypothesis III states that, in secondary schools, no evaluation effect will be observed as a function of subject taught. To test this hypothesis, we examined mean ratings of teachers of mathematics, science, language arts, foreign language, and social studies. These were compared to mean ratings of all teachers and to other classes. Table IIIa presents this information, in numerical form. Clearly, teachers of mathematics, foreign language, and language arts exceeded the general mean on every function in both years, while science and social studies teachers were at or below the mean both years. Table IIIb indicates that social studies teachers, as a group, failed to attain the general mean on any function either year, while mathematics, foreign language, and language arts teachers exceeded the general mean on each function in both years. As Table IIIc makes clear, the means for each function improved in 1989 when compared with the 1988 means. However, these data may suggest that teachers of more structured courses (math, foreign language) generally fare better than do teachers of less structured courses. While this could explain ratings for Function 3 (Instructional Presentation), it does not explain ratings for, say, time management or student

TABLE IIIa
Mean Ratings by Function for Teachers by Subject Area

1988

FUNCTION	All	Lang. Arts	Math	Science	Soc. Stud.	Foreign Lang.
1	4.56	4.62	4.72	4.56	4.32	4.65
2	4.56	4.57	4.68	4.44	4.44	4.58
3	4.51	4.65	4.77	4.61	4.34	4.75
4	4.54	4.56	4.76	4.46	4.32	4.81
5	4.51	4.57	4.78	4.51	4.32	4.83
	(N=5835)	(N=263)	(N=213)	(N=187)	(N=179)	(N=80)

1989

FUNCTION	All	Lang. Arts	Math	Science	Soc. Stud.	Foreign Lang.
1	4.71	4.73	4.80	4.74	4.56	4.80
2	4.71	4.72	4.78	4.61	4.66	4.75
3	4.65	4.82	4.80	4.76	4.54	4.79
4	4.66	4.67	4.88	4.54	4.42	4.82
5	4.62	4.66	4.84	4.56	4.43	4.81
	(N=5915)	(N=272)	(N=211)	(N=160)	(N=177)	(N=84)

TABLE IIIb
Mean Ratings By Subject Area Compared to General Mean Rating

Func.	88	89	88	89	88	89	88	89	88	89
	Math		Foreign Language		Language Arts		Science		Social Studies	
1	+	+	+	+	+	+	-	+	-	-
2	+	+	+	+	+	+	-	-	-	-
3	+	+	+	+	+	+	+	+	-	-
4	+	+	+	+	+	+	-	-	-	-
5	+	+	+	+	+	+	-	-	-	-

TABLE IIIc
Change of Mean Ratings from 1988 to 1989
By Subject Area

Func.	MATH	SCIENCE	FOREIGN LANGUAGE	LANGUAGE ARTS	Social Studies
1	+	+	+	+	+
2	+	+	+	+	+
3	+	+	+	+	+
4	+	+	+	+	+
5	+	+	+	+	+

behavior management. However, it can be argued that more structured content courses (skills) require less external control than do more loosely defined courses. Is there something inherent to the way science and social studies are taught that augurs poorly for these teachers when evaluated on a generic skills measure? We cannot, given these data, answer this question which is, essentially, Hypothesis III.

Discussion

The NCTPAS was predicated on the existence of generic teaching skills that are appropriate in any teaching assignment. A careful review of the research base had isolated 28 such skills that were tested in a variety of courses and classes. As the researchers reported, then, however, none of the skills had been reported in every course and every grade (The Group, 1985). Before undertaking the current study, staff of DPI had reported mean scores for all teachers in the sample, without attending to differences of assignment. Generally speaking, performance means had increased over the four-year period.

However members of specific teacher classes had registered uncertainty about the basic premise almost from the beginning. This uncertainty sometimes was expressed as a vague uneasiness that "an evaluator who knows nothing about _____(French, special education, kindergarten, etc.) will be unable to evaluate my teaching fairly". This uneasiness led at least one researcher to study the evaluations of vocational educators and their perceptions of their own skills (Jewell, 1989.) Despite the fact that his study group clearly demonstrated acceptance of the criteria as

appropriate, the research suggests that in "hands-on or off-site teaching situations" a different evaluation might be necessary. (Jewell, p. 12).

A somewhat different approach to the issue of applying the outcomes of "generic skills" research has been taken by special educators. David B. Ryan, a special educator employed by one of the CDP districts as an evaluator of teachers, developed a comprehensive report that reviewed the special education literature and related the NCTPAS practices to this literature base (Ryan, n.d.) In this effort, he continued the work of Bickel and Bickel (1986) who reviewed the effective schools, effective classrooms and effective instruction literature to find connections to special education, and of Morsink, Soar, Soar, and Thomas (1986). In an empirical test of application of a generic teacher evaluation model in special education, Algozzine, Morsink, and Algozzine (1986) found that 17 of the 22 criteria in that model were appropriate for evaluating special educators.

The data presented in this study suggest that a generic skills approach to evaluation does not negatively impact teachers of classes or grades that differ in significant ways from the "normal" classroom. Teachers' evaluations, in other words, reflected what they did rather than where they did it or to whom they did it. The same general trends found for "all teachers" seem to be upheld when any sub-set is examined individually.

The most interesting question revealed in this paper may indeed be the issue of "skills" courses as opposed to "knowledge" courses. While this distinction is crude, what we have in mind are

courses designed to teach hierarchically-related sequences of knowledge (mathematics) as opposed to courses in which content is less clearly sequenced and/or results in less clear demonstration of knowledge. Rosenshine (in Peterson, 1979) distinguishes these two genres and the data we report for secondary school teachers could be taken to support this notion.

Conclusion

During the past six years, educators in North Carolina have made significant progress in the area of evaluating teaching skill. The contribution of research efforts for establishing a knowledge base cannot be overstated. During the 1960's and 1970's, a disparate group of researchers, supported largely by federal dollars, began the effort of untangling the complex behaviors teacher engage in to study the effects on students' learning. By this time, there can be little question about the essential skills of teaching: managing time, managing student behavior, instruction, monitoring, and feedback. Indeed, in twelve states that employ state-wide teacher evaluation systems, these skills are at the heart of the evaluative criteria (French, Holdzkom, and Kuligowski, 1990).

Indeed, while "teacher evaluation" has been at the core of much contentious discussion, it is interesting that the criteria for evaluation are less often the issue than the who, how, and why questions. Bacharach and his colleagues attack evaluation because of its potential for establishing accountability (Bachrach, 1990). Some argue that evaluation could be conducted for performance improvement, but only in a "collegial" manner (Glatthorn, 1990).

Darling-Hammond echoes this position when she calls for peer evaluation and attacks systems that call for principals to be primary evaluators. The use of simplistic check-lists or "bubble-sheets" has been attacked, appropriately, by many (Wise, et al, 1984). However, the criteria (the generic skills of teaching) are less often attacked and very few groups offer acceptable alternatives, although Shulman (1986) and his colleagues represent an obvious exception.

Beginning in 1985, the North Carolina Department of Public Instruction systematically studied the implementation of a system of evaluation of generic teaching skills. We have established several things:

1. Principals and other supervisors can be trained to recognize and evaluate generic teaching skills.
2. Evaluators' skills improve over time, especially when additional training is made available.
3. Teachers' generic teaching skills improve, especially when they are given specific feedback and additional training.
4. A generic skills evaluation system does not discriminate unfairly against any class of teachers.
5. Students' basic skills achievement rises when taught by more skilled teachers.
6. Teachers and evaluators agree that the criteria and processes of the evaluation system are reasonable and fair.

However, it is obvious that a generic skills approach to

performance evaluation has a finite ability to improve, or even account for, teaching skill, to say nothing of the link to learning. A generic skills approach is necessary but may not be sufficient to evaluate teaching. What, then, remains to be done?

Clearly the next steps will involve the study of content-specific pedogogy, to borrow Shulman and Berliner's term (Shulman, 1986; Berliner, 1986). How this is to be done is much less clear. The very limited results of Shulman's work are instructive but appear not to be generalizable. Federal funds to support research in this area are much less than were available to support the process-product research (Cross, 1990). A common research agenda seems to exist only in a distant future. Efforts by the National Board for Professional Standards in Teaching (Baratz-Snowden, 1990) may help, as could efforts by the subject-area associations (e.g., NCTE, NSTA, etc.) Indeed, National Council of Teachers of Mathematics has made a start in this direction. However, until such research and development activities are completed, we, as a profession, are unlikely to be able to move beyond evaluation of essential skills.

BIBLIOGRAPHY

Algozzine, Bob; Morsink, Catherine V.; and Algozzine, Kate M. 1988. "What's happening in self-contained special education classrooms?" Exceptional Children, v.55 #3, pp 259-265.

Barcharach, Samuel B.; Conley, Sharon C.; and Shedd, Joseph D. "Evaluating Teachers for Career Awards and Merit Pay" in Millman, Jason and Darling-Hammond, Linda. (eds). 1990. The New Handbook of Teacher Evaluation. Newbury Park, C.A.: Sage Publications.

Baratz-Snowden, Joan. 1990. The NBPTS begins its research and development program. Educational Researcher, v.19, #5, pp 19-24.

Berliner, David. 1986. In Pursuit of the Expert Pedagogue. Educational Researcher. v.15 (7) pp 5-13.

Bickel, William E, and Bickel, Donna DiPrima. 1986. "Effective Schools, classrooms, and instruction: Implications for Special education Exceptional Children., v.52, #6 pp 489-500.

Brandt, Richard M. 1990. Career Ladders for Today's Teachers. Albany, N.Y.: SUNY Press.

Brandt, Richard M., Duke, Daniel L., French, Russell L., and Iwanicki, Edward F. 1988. A Review With Recommendations of the North Carolina Teacher Performance Appraisal Instrument. Raleigh, N.C.: North Carolina General Assembly (Joint Legislative Commission on Government Operations.) p 2.

Cross, Christopher T. 1990. National goals: Four priorities for educational researchers. Educational Researcher, v.19, #8, pp 21-24.

Darling-Hammond, Linda. 1989. Teacher Professionalism and Accountability. The Education Digest. September.

French, Russell, Holdzkom, David, and Kuligowski, Barbara. 1990. Teacher Evaluation in SREB States: Stage I -- Analysis and Comparison of Evaluation Systems. Atlanta, G.A.: Southern Regional Education Board.

Furtwengler, Carol. (Research and Service Institute). 1988. Evaluation of N.C.'s School Career Development Pilot Program. Raleigh, N.C.: Joint Legislative Commission on Government Operations.

Glatthorn, Allan A. 1990. Supervisory leadership: Introduction to Instructional Supervision. Glenview, I.L.: Scott Foresman.

The Group for Effective Teaching. 1983. Teaching Effectiveness Evaluation Project: Final Report. Raleigh, N.C.: Quality Assurance Program (NCDPI.)

The Group for Effective Teaching. 1985(a). Carolina Teaching Performance Assessment System: Observer's Manual. Raleigh, N.C.: Quality Assurance Program (NCDPI.)

BEST COPY AVAILABLE

The Group for Effective Teaching. 1985(b). Carolina Performance Assessment System: Results of a Limited Field test. (Final Report). Raleigh, N.C.: Department of Public Instruction.

Holdzkom, David. 1987. "Appraising Teacher Performance in North Carolina. Educational Leadership. v.44(f), pp 40-44.

Holdzkom, David; Kuligowski, Barbara; Stacey, Dennis C. 1989. A Longitudinal Study of Teacher Performance Evaluation. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, C.A.

Holdzkom, David, Kuligowski, Barbara, and Stacey, Dennis C. 1990. Better Teaching for Better Learning: Student Achievement Results in a 4-Year Pilot of the N.C.C.D.P. Paper presented at the annual meeting of the American Educational Research Association, Boston, M.A.

Jewell, Larry R. 1989. Assessing Teaching Performance of Vocational Education Teachers in the Central Region of North Carolina. Paper presented to the Region III Vocational Education Research and Development Project Advisory Committee, Raleigh, N.C.

Movsink, Catherine V.; Soar, Robert M.; Soar Ruth M.; and Thomas, Roberta. 1986. "Research on teaching: Opening the door to special education classrooms." Exceptional Children, v.53, #1, pp 32-40.

North Carolina Department of Public Instruction. 1987. Performance Appraisal: Cornerstone of Career Development Plan. Raleigh, N.C.: State Board of Education.

Rosenshine, Barak. 1979. "Content, Time, and Direct Teaching" in Peterson, P.L. and Wolberg, H.J. (eds), Research on Teaching: Concepts, Findings, and Implications. Berkeley, C.A.: McCutcheon Publishing.

Ryan, David B. (n.d.) Evaluating Special Education Teachers. Wilmington, N.C.: New Hanover County Schools.

Shulman, Lee S. 1986. "Those Who Understand: Knowledge Growth in Teaching." Educational Researcher. v. 15 (2), pp 4-14.

Stacey, Dennis C., Holdzkom, David, Kuligowski, Barbara. 1989. "Effectiveness of the North Carolina Teacher Performance Appraisal System." Journal of Personnel Evaluation in Education. 3: 79-106.

Stodolsky, Susan S. 1984 "Teacher evaluation: The limits of looking." Educational Researcher, v. 13 (9) pp 11-18.

Wise, Arthur E., Darling-Hammond, Linda, McLaughlin, Milbrey W., and Bernstein, Harriet T. 1984. Teacher Evaluation: A Study of Effective Practices. Santa Monica: The Rand Corporation.